A COMPARATIVE STUDY OF NONLINEAR VIDEO RATE CONTROL TECHNIQUES: NEURAL NETWORKS AND FUZZY LOGIC

Yoo-Sok Saw†, Peter M. Grant‡ and John M. Hannah‡
†Centre for Communications Research, Dept. of Electrical Engineering, Univ. of Bristol, Bristol, BS8 1UB, UK.
Tel: +44 117 954 5125; Fax +44 117 954 5206
e-mail: Yoo-Sok.Saw@bris.ac.uk
‡Dept of Electrical Engineering, University of Edinburgh, Edinburgh, EH9 3JL, UK.
Tel: +44 131 650 5655; Fax: +44 131 650 6554
e-mail: Peter.Grant@ee.ed.ac.uk

ABSTRACT

Data rate management of compressed digital video has been treated mainly from the teletraffic control point of view, i.e. by modelling traffic or controlling congestion via network protocols. Relatively less attention has been focused on video rate management in the source coding side. In this paper we consider that it is more efficient and less costly to control video rate at the video source than handling network congestion due to an extremely large quantity of incoming variable bit rate (VBR) video traffic. Thus this paper investigates effective rate control algorithms for video encoders. Considering the non-stationary nature of video rate originated from scene variations (i.e. the wide band nature of digital video), we adopted and compared the performance of two nonlinear approaches; radial basis function (RBF) estimation using a neural network-based approach and fuzzy logic control as a nonlinear feedback control.

1. INTRODUCTION

Video rate control is concerned with regulating video data rate in conjunction with video quality when video is encoded by statistically based compression schemes such as MPEG (Moving Picture Experts Group) standards. There is a trade-off relationship between data rate of compressed video and resulting quality.

1.1. Background of video rate control

An effective rate control algorithm becomes more demanding when the video contains rapid motion or frequent scene changes mixed with less scene variations. This wide band nature of digital video is a major cause of abrupt variations in video rate. In this paper attention is focused mainly on video with rapid scene variation, Figs.1 and 2 ("Adverts", TV advertisements and movie "Star Wars"), which cannot easily be accommodated in conventional schemes such as TM5 and linear prediction-based methods. As an alternative, we introduce a priori knowledge (i.e. *scene change features* to represent variation in visual information) to improve the rate control performance of the proposed rate control algorithms employed in the MPEG video encoder.

Although the MPEG2 evaluation model, TM5 specifies details of its rate control process, the scheme is based on a linear prediction technique and is known to be inappropriate for video with large scene changes. TM5 is presented as a performance reference for comparison purposes in this paper.

Figure 1. Selected frames from the "Adverts" sequence.

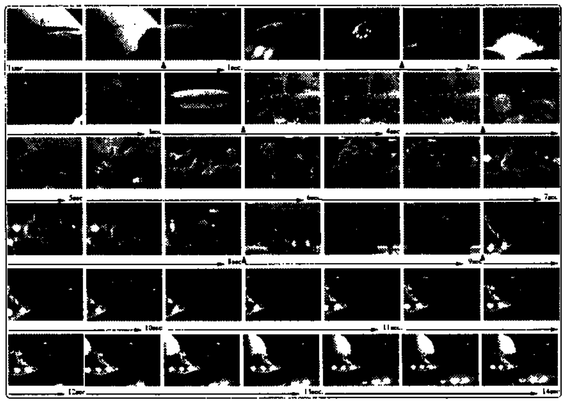


Figure 2. Selected frames from the "Starwars" sequence.

1.2. Feed-forward video rate control: a neural network approach

One of novel architectures applied in this paper is the nonlinear predictive video rate estimator using the RBF network. The RBF network is often classified as a special form of neural network with a smaller number of hidden layers. It also has an advantage in the aspects of implementation, as it is a parallel structure. The RBF network is known to have better estimation performance than linear predictors for non-stationary signals.

We employed the RBF network in a feed-forward predictive scheme, which exploits short-term correlation of video and improves the estimation performance. Although one can achieve a certain level of performance by using linear techniques, the performance can be further improved if nonlinear predictive technique is used. In this feed-forward scheme, the scene change features are used as the input for the RBF video rate estimator. They represent the 1st and 2nd-order statistical measures: intra-frame variance, inter-frame variance plus integer picture type (I, P and B) values.

1.3. Feedback video rate control: a fuzzy logic approach

The buffer-based video rate control is fundamentally a feedback control in which the buffer occupancy is translated into a quantisation step size. As an improved feedback approach, we investigated a fuzzy-logic based video rate control technique. It is considered that conventional fuzzy-logic based control (FLC) does not effectively control the two variables (video data rate and video quality) which are in a trade-off relationship. The primary reason for this is that it is not easy to effectively project the control variables onto fuzzy rules due to the contradiction between these two variables on a rate-distortion theoretic basis.

2. FEED-FORWARD PREDICTIVE VIDEO RATE ESTIMATION FOR MPEG

The algorithm of the feed-forward RBF-network-based scheme, Fig.3, is described in detail in [1]. In this scheme two of the three scene change features are framewise variances, $var\_org(k)$ and $var\_dif(k)$, which represent the variance of the input picture and the variance between the input picture and the previous picture, respectively, where k represents picture time index. The other scene change feature is picture type, $ptype(k)$, and a single value exists for the corresponding picture type, thus it forms a cyclic time series as k proceeds as the picture type repeats, e.g. IBBPBBPB...

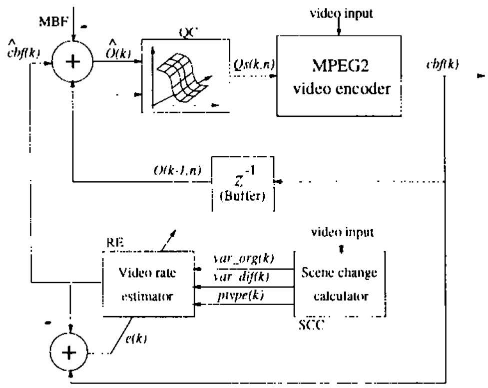


Figure 3. The structure of the feed-forward network for MPEG rate control.

The RBF network comprises the rate estimator (RE). The estimated video rate $\widehat{cbf}(k)$ is used to derive an appropriate quantiser step size in association with the current buffer occupancy $O(k-1,n)$. QC represents the nonlinear quantiser control surface: sigmoidal or unimodal [1]. MBF stands for the number of mean bits per frame.

2.1. Training and Optimising the RBF Video Rate Estimator

In consideration of the configuration of the MPEG2 video encoder, the number of RBF centres is set to 9. The input consists of 3 scene change features, each of which has 3 taps. The number of time delays is equal to the number of B pictures between P pictures. It is assumed that the cyclic repetition of the video rate is determined by the number of B pictures, and the correlation of the video rate varies with an interval of two pictures. Different numbers of centres - up to 50 - were also simulated and their performance was compared in normalised mean square error (NMSE), as shown in Fig. 4 and Table 1. We used realistic video sequences including two other sequences, "Cascaded" and "JFK", to see the effect of the RBF network estimator. Three standard sequences, "Miss America", "Football" and "Susie", comprise "Cascaded". "JFK" is an edited version of the movie "JFK" with rapid scene variation. For all four video sequences tested, the 6 to 9 centres appeared to exhibit very similar performance to cases of the larger number of centres. This signifies that the number of RBF centres may be set in conjunction with the number of the B pictures. This implies that the system complexity of a RBF-network-based rate control can be dramatically reduced by selecting an appropriate number of centres from video encoding parameters.

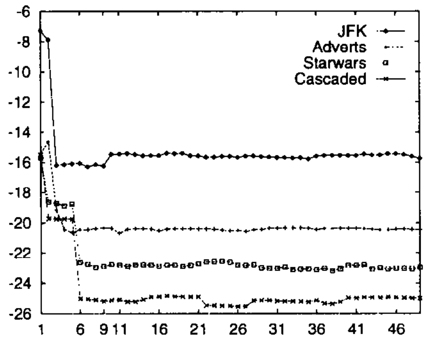


Figure 4. Variation in NMSE profile with the number of centres.

	MSE[dB]	*var_org(k)*	*var_dif(k)*
Cascaded	-24.49	1224.8	579.6
Starwars	-22.44	1671.2	713.4
Adverts	-20.67	2888.5	1143.9
JFK	-15.33	3329.9	1914.7

Table 1. Mean NMSE and mean variances of video sequences.

3. FUZZY-LOGIC BASED VIDEO RATE CONTROL: A FEEDBACK APPROACH

3.1. A Basic FLC model for video rate control

Although an analogy can be found in other fuzzy logic-based industrial applications, video rate control is not just a matter of controlling the level of liquid reservoir which is treated as a typical FLC application. In order to improve the FLC performance for video rate control, we introduced adaptive scaling factors which vary, depending on the scene change features. First, we examined the conventional FLC where the number of fuzzy variables is one, and its performance was evaluated in various settings of the fuzzy control parameters. As an adaptive algorithm, a FLC scheme with feed-forward scaling factors was designed and its performance was compared to the conventional FLC.

Recently, a few leading researchers have applied the FLC to video sequence coding standards. The techniques aim to improve the video rate control performance for JPEG [2, 3] and H.261 [4] by adaptively controlling the quantiser and the buffer occupancy. The fuzzy logic-based control techniques used in these researches have the same technical base in that they appeared to follow a series of common processes; fuzzification, decision making and defuzzification.

Figure 5 shows the configuration of such a FLC-based video rate control (FLC-R) which takes the buffer occupancy as its only input, with its parameter settings shown in Fig.6. The control input, Go, which is used as the quantisation step size, $Qs(n)$, is the input to the encoder with the macro block time index n. The process of the FLC begins with calculating the error value, $e(n)$, and its differential error value, $d(n)$, which is the difference between the current error value, $e(n)$, and the previous error value, $e(n-1)$. Its final output is obtained by a series of arithmetic operations, e.g. the centre of gravity method [5]. All scaling factors (ge, gd and go) are constants which are generally tuned by expert knowledge.

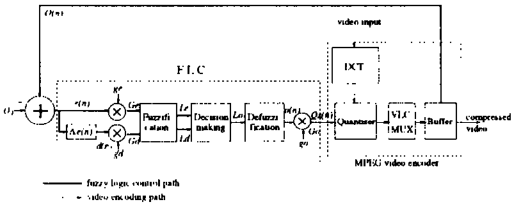
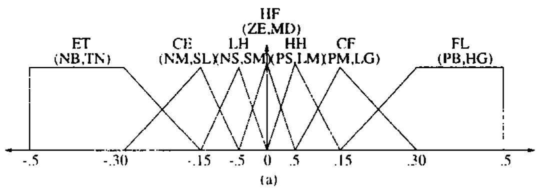
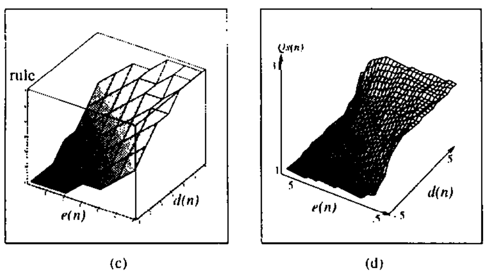


Figure 5. Configuration of the FLC-based rate control.



Ge \ Gd	NB	NM	NS	ZE	PS	PM	PB
ET	TN	TN	TN	SL	SL	SM	MD
CE	TN	TN	SL	SL	SM	MD	LM
LH	TN	SL	SL	SM	MD	LM	LG
HF	SL	SL	SM	MD	LM	LG	LG
HH	SL	SM	MD	LM	LG	LG	HG
CF	SM	MD	LM	LG	LG	HG	HG
FL	MD	LM	LG	LG	HG	HG	HG

(b)



(c) (d)

Figure 6. Fuzzy logic-based parameters for video rate control. (a) membership function, (b) FAM, (c) 3-dimensional representation of the FAM, (d) the resulting control surface.

The membership function can take different shapes, different inter-rule spacing, etc. It is well known that using non-triangular shapes does not provide substantial difference in the performance, hence, the triangular shape is used here. The formation can be asymmetrical since the positive section of the $e(n)$ value can have different significance from the negative section. In video rate control, however, both sections here are assumed to have unbiased linguistic interpretation. Thus, the membership functions shown in Fig. 6(a) are symmetrical with respect to the centre value 0. Fig. 6(b) shows the configuration of the rule base to determine the dynamic behaviour of the FLC system. The 3-dimensional representation of FAM, Fig. 6 (c), reflects the dynamic property of an organisation of rules and membership functions. Fig. 6(d) shows the resulting control surface for the FAM configuration of Fig. 6(b) and (c) with the two input variables, $d(n)$ and $e(n)$.

3.2. FLC with adaptive scaling factors

Scene change features provide vital information about the resulting number of bits for an incoming picture in advance of encoding it. Here, they are incorporated with the non-adaptive scaling factors (ge, gd and go) of FLC-R, as shown in Fig. 7, in order to adaptively scale inputs of the fuzzification process, $e(n)$ and $d(n)$. This configuration forms the adaptive FLC scheme (FLC-FS). The three previous scene change features are supplied to the scaling factor calculation block (Scaling function) which generates time varying scaling factor values, $ge(k)$ and $gd(k)$. The remainder of the FLC-FS operation is the same as FLC-R depicted in Fig. 5. In the mapping function the equation maps the scene change features to the scale factors as follows:

$$ge(k) = gd(k) = \frac{\log_{10} var\_org(k)}{\log_{10} var\_dif(k)} \times ptype(k) \quad (1)$$

In this scheme, the scaling factors vary depending on the scene changes features so that the video rate can be controlled effectively for the next video image. This enhances the performance of the FLC-based rate control scheme.

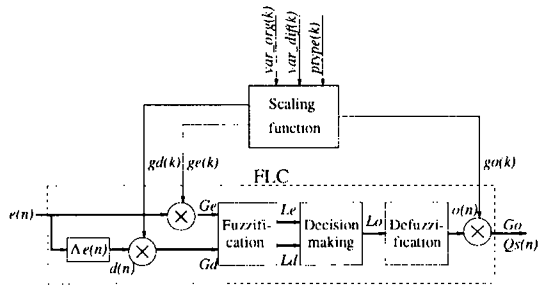


Figure 7. Configuration of the FLC-FS.

4. SIMULATION STUDIES

The performance of the feed-forward rate estimators is first evaluated and FLC schemes will be compared with rate estimators on the test bed of the MPEG2 software encoder. The new configuration of the MPEG2 encoder was verified with a variety of different settings of encoding parameters. However, we only present representative simulation results for the "Starwars" sequence.

The MPEG2 video encoder based on TM5 was set to operate at a channel rate of 1024 kbits/s and a frame rate at 30 frames/s. It has a buffer with the size of twice the MBF. This size of buffer is specified for low delay mode codings in bi-directional communications. The unimodal function was used as the quantiser control surface since it possesses better performance for video rate control than the sigmoidal function. For the feed-forward rate control scheme, three techniques are evaluated: the two rate estimator-based schemes, RLS (recursive linear estimation technique) and RBF.

Table 2 shows the mean and standard deviation for each of the performance measure over the whole sequence. NFVR in the middle column stands for normalised fluctuation of the video rate which represents variation of $cbf()$. It is expressed in the following equation:

$$\mathrm{NFVR} = \frac{\sigma}{1+\sigma}, \qquad \sigma^2 = E\left[\left(\frac{cbf(k)}{\mathrm{MBF}} - 1\right)^2\right] \tag{2}$$

where $\frac{cbf(k)}{\mathrm{MBF}}$ represents instantaneous fluctuation.

TM5 exhibits the worst occupancy performance, often reaching buffer overflow on some sequences such as "JFK", with more rapid movement. The schemes based on predictors, RLS and RBF, show better performance than TM5 in both bit rate and PSNR. RBF appeared to be capable of maintaining the occupancy lower with a smaller standard deviation in comparison to RLS. RBF also exhibits a similar PSNR value to RLS. This result implies that the nonlinear rate estimator, RBF-network, works more effectively for non-stationary video with many scene changes and rapid motion without further degradation in video quality.

The basic FLC model (FLC-R) possesses a considerable flexibility to change the fuzzy control parameters. The scaling factor for the output $go(k)$ was set to 1. Pictures, which entail a large number of bits, will be given much stronger control action. When they are I or P pictures, the distortion caused by rate control will affect the next coming pictures. For this reason PSNR remains low when scaling factors are large.

The FLC model assisted by the feed-forward scaling factors (FLC-FS) was assessed with respect to FLC-R. While FLC-R is superior in controlling the video rate or the occupancy to FLC-FS, the latter shows wider variations since its scaling factors change depending on scene change. Hence, profiles of the occupancy and video rate may vary dramatically depending on the scene change. The video rate profile exhibits similarly changing patterns.

FLC-R consists of fuzzy control rules based only on the occupancy-related rules. Therefore, it appeared to be powerful in controlling the occupancy. However, it does not take account of the quality, and it shows lower figures in PSNR. For FLC-FS, the scaling factors are allowed to change within a 1-to-8 range depending on the scene change features as specified Eqn. 1. In FLC-FS, the scaling factors $ge(k)$ and $gd(k)$ scale up the error signal, $e(n)$, Fig. 7, to adaptively change the actual input of $e(n)$. The scaling factors are generally smaller than 8, and the resulting performance of the occupancy control appear to be inferior to FLC-R. Accordingly, FLC-FS shows more fluctuating profiles. However, the feed-forward scaling factors ($ge(k)$ and $gd(k)$) improve the video quality by making the most of the occupancy margin below 50%. Therefore, it is concluded that FLC-FS performs better than FLC-R in terms of PSNR achieved as well as the occupancy.

Comparing the FLC-FS scheme with the RBF estimator-based scheme, the performance shows no noticeable difference as shown in Table 2. Thus, comparison between these two schemes should be considered from a different point of view: performance for particular scene change types, implementational complexity, etc.

Starwars	Occupancy(%)		Coded bits / frame		PSNR (dB)	
	mean(max.)	std.dev.	NFVR	std.dev.	mean	std.dev.
TM5	41 (75)	10.8	0.285	13704	33.27	2.69
RLS	18 (38)	4.5	0.127	4519	33.89	2.54
FLC-R	27 (32)	1.5	0.086	1321	33.40	2.59
FLC-FS	19 (42)	5.3	0.146	5058	33.86	2.55
RBF	12 (34)	4.3	0.124	4288	33.87	2.54

Table 2. Performance comparison between FLC-R and FLC-FS ("Starwars").

5. CONCLUSIONS

Improvements on video rate control have been achieved using nonlinear approaches: RBF-network and fuzzy logic control. These two approaches have different structures and properties, thus, performance comparison is not straight forward. However, a comparison can be made in terms of feasibility. The fuzzy logic approach is considered to be computationally simpler and less complicated, which is critical to implementation. However, the simulation results show that the RBF-network scheme does exhibit a lower mean buffer occupancy than the fuzzy logic-base control schemes. This feature is advantageous for low-delay applications since shorter delay can be achieved by maintaining low buffer occupancy. In the RBF-network-based scheme, supplementary processing is required on the estimated video rate information to derive a corresponding quantisation step size to achieve the predictive rate control function. On the other hand, the fuzzy logic-based control can accomplish the equivalent task by using a set of rules, a pair of control variables and adaptive scaling factors. This configuration is generally simpler than the approach for predictive scheme. Hence, as the performance is very similar, the fuzzy logic approach is more promising in terms of system complexity than the RBF-network-based scheme.

6. ACKNOWLEDGEMENTS

Yoo-Sok Saw acknowledges the support of *The British Council* and *LG Information and Communications Ltd., Korea*.

REFERENCES

[1] Y. Saw, P. M. Grant, J. M. Hannah, and B. Mulgrew, "Nonlinear predictive rate control for constant bit rate mpeg video coders," in *ICASSP'97*, vol. 4, pp. 2641–2644, IEEE, 21-24, April, 1997.

[2] C. Wu and A. Sung, "Application of fuzzy controller to JPEG," *IEE Electronics Letters*, vol. 30, no. 17, pp. 1375 1376, 1994.

[3] S. H. Supangkat and K. Murakami, "Quantity control for JPEG image data compression using fuzzy logic algorithm," *IEEE Trans. Consumer Electronics*, vol. 41, no. 1, pp. 42 - 48, 1995.

[4] A. Leone, A. Bellini, and R. Guerrieri, "An H.261-compatible fuzzy-controlled coder for videophone sequences," in *3rd Int. Conf. on Fuzzy Systems*, vol. 1, pp. 244– 248, IEEE, 26 - 29 June, 1994, Orlando, Florida.

[5] J.-S. R. Jang and C.-T. Sun, "Neuro-fuzzy modeling and control," *IEEE Proceedings*, vol. 83, no. 3, pp. 378 – 406, 1995.